Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The N-terminal substrate-recognition domain of a LonC protease exhibits structural and functional similarity to cytosolic chaperones

Jhen-Kai Li,[a] Jiahn-Haur Liao,[b] Hongchun Li,[c,d] Chiao-I Kuo,[b] Kai-Fa Huang,[b] Lee-Wei Yang,[d] Shih-Hsiung Wu[a,b] and Chung-I Chang[a,b]*

[a]Institute of Biochemical Sciences, College of Life Science, National Taiwan University, Taipei 10617, Taiwan, [b]Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan, [c]Department of Chemistry, College of Chemistry and Chemical Engineering, and the Key Laboratory for Chemical Biology of Fujian Province, Xiamen University, Xiamen, People's Republic of China, and [d]Institute of Bioinformatics and Structural Biology, National Tsing-Hua University, Hsinchu, Taiwan

Correspondence e-mail: chungi@gate.sinica.edu.tw

Received 9 April 2013
Accepted 31 May 2013

PDB Reference: N-terminal domain of MtaLonC, 4fwv

The Lon protease is ubiquitous in nature. Its proteolytic activity is associated with diverse cellular functions ranging from maintaining proteostasis under normal and stress conditions to regulating cell metabolism. Although Lon was originally identified as an ATP-dependent protease with fused AAA+ (ATPases associated with diverse cellular activities) and protease domains, analyses have recently identified LonC as a class of Lon-like proteases with no intrinsic ATPase activity. In contrast to the canonical ATP-dependent Lon present in eukaryotic organelles and prokaryotes, LonC contains an AAA-like domain that lacks the conserved ATPase motifs. Moreover, the LonC AAA-like domain is inserted with a large domain predicted to be largely α-helical; intriguingly, this unique Lon-insertion domain (LID) was disordered in the recently determined full-length crystal structure of *Meiothermus taiwanensis* LonC (MtaLonC). Here, the crystal structure of the N-terminal AAA-like α/β subdomain of MtaLonC containing an intact LID, which forms a large α-helical hairpin protruding from the AAA-like domain, is reported. The structure of the LID is remarkably similar to the tentacle-like prong of the periplasmic chaperone Skp. It is shown that the LID of LonC is involved both in Skp-like chaperone activity and in recognition of unfolded protein substrates. The structure allows the construction of a complete model of LonC with six helical hairpin extensions defining a basket-like structure atop the AAA ring and encircling the entry portal to the barrel-like degradation chamber of Lon.

1. Introduction

Chambered proteases play an important role in degrading misfolded or abnormal proteins, which are otherwise prone to form toxic aggregates inside cells (Wickner *et al.*, 1999; Selkoe, 2003). These intracellular multi-component/subunit proteases often form an enclosed proteolytic chamber to ensure that access to the chamber is limited to protein substrates with certain exposed structural or sequence features that can be recognized by a substrate-recognition component of the protease outside the chamber. In most cases, the substrates are then unfolded and fed into the chamber by the oligomeric ATPase component of the protease powered by ATP hydrolysis (Sauer & Baker, 2011; Gur *et al.*, 2012).

Lon was the first ATP-dependent protease to be characterized and was later found to form a ubiquitous family of proteases that contain in a single polypeptide chain an AAA+ (ATPases associated with diverse cellular activities) module and a unique C-terminal protease domain with a conserved serine–lysine catalytic dyad (Goldberg *et al.*, 1994; Rotanova *et*

al., 2006). Lon proteases play various important roles in cells, ranging from maintaining proteostasis to metabolic regulation by degrading specific protein substrates (Granot *et al.*, 2007; Mizusawa & Gottesman, 1983). Based on domain organization, Lon proteases can be classified into three types, each of which contains a type-specific Lon-insertion domain (LID). For example, members of the LonA type found in all bacterial species and in eukaryotic organelles, such as mitochondria, chloroplasts and peroxisomes, have a LID (known as LAN) at the N-terminus of the AAA+ module. LonB members, which exist exclusively in archaea, have a LID located within the AAA+ module containing two transmembrane segments (Rotanova *et al.*, 2006). The recently characterized ATP-independent LonC proteases possess a large LID (~200 amino acids), predicted to be mainly α-helical, embedded within the AAA-like α/β-subdomain (Fig. 1*a*; Liao *et al.*, 2013).

No full-length structure is available for any Lon protease. Previously, the structure of a truncated LonB protease without the LID has been reported (Cha *et al.*, 2010). This structure showed a sequestered hexameric barrel-like degradation chamber assembled from the fused rings of the AAA+ and protease domains. Recently, the structures of a full-length LonC protease and its complexes with covalent inhibitors have been determined (Liao *et al.*, 2013). Although the overall structure of the LonC monomer is similar to that of LonB, the assembled LonC chamber contains two open (~13 Å) axial pores rather than being completely occluded as in LonB. Despite these advances, neither of the two structures revealed the architectural features of the LID located atop the Lon chamber, as it was either not included in the crystal construct (Cha *et al.*, 2010) or disordered in the crystal (Liao *et al.*, 2013). As the AAA+ ring forms the proposed substrate-entry portal to the interior chamber, the LID may play a pivotal role in substrate interaction. Indeed, mutational and biochemical analyses have suggested that the N-terminal LID/LAN of LonA is essential for substrate recognition, which is also intricately involved in allosteric regulation of the ATPase and proteolytic activities of Lon (Gur, 2013; Cheng *et al.*, 2012; Vasilyeva *et al.*, 2002; Ebel *et al.*, 1999; Roudiak & Shrader, 1998). Crystallographic results on isolated N-terminal fragments of LonA have shown that this region consists of a globular domain and an extended helical element (Li *et al.*, 2005, 2010; Duman & Löwe, 2010). However, none of these structures offer a complete view of an intact LID. Here, we have determined the crystal structure of the N-terminal region of a LonC protease containing the AAA+ α/β subdomain with an intact LID, which provides functional insight into its role in substrate engagement of the Lon protease.

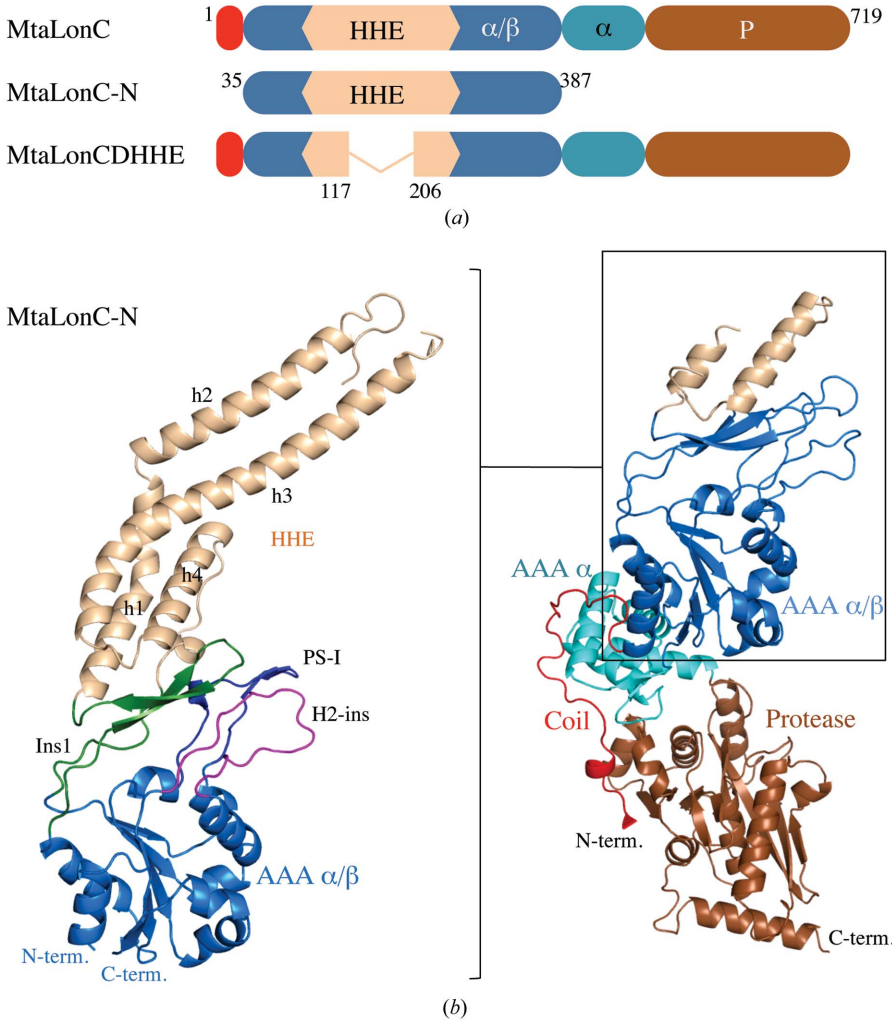

Figure 1
Overall structure of MtaLonC-N. (*a*) Domain organizations of wild-type MtaLonC, MtaLonC-N and MtaLonCΔHHE used in the present study (see text). (*b*) Ribbon diagrams of MtaLonC-N (left) and apo MtaLonC (right; Liao *et al.*, 2013) showing the domain arrangements in a similar colouring scheme as in (*a*). Also indicated are the N- and C-termini and the three classified AAA inserts (Ins1, PS-I and H2-ins; Iyer *et al.*, 2004), which are shown in different colours. The bracket denotes the portion of apo MtaLonC corresponding to MtaLonC-N, which contains the helical hairpin extension (HHE; residues 96–252), shown as a wheat-coloured ribbon.

2. Materials and methods

2.1. Molecular cloning, protein expression and purification

The cDNA fragment corresponding to residues 35–387 of the N-terminal domain of *Meiothermus taiwanensis* LonC (MtaLonC-N) was cloned into pET28a (Novagen) pre-modified by replacing its thrombin-cleavage site (Leu-Val-Pro-Arg-Gly-Ser), which is located after the 6×His coding sequence, with a *Tobacco etch virus*

Table 1
Data-collection, phasing and refinement statistics.

Values in parentheses are for the highest resolution shell. ND, not determined.

(*a*) Data collection.

Crystal (space group *R*32)	SeMet-MtaLonC-N, crystal 1	SeMet-MtaLonC-N, crystal 2	MtaLonC-N, crystal 3 (dehydrated)
Wavelength (Å)	0.9789 [peak; λ_1]	0.9639 [remote; λ_2]	0.9999 [native]
Resolution range (Å)	50–3.40 (3.52–3.40)	50–3.13 (3.21–3.13)	30–2.40 (2.49–2.40)
Unit-cell parameters (Å)			
$a = b$	147.6	147.3	147.7
c	165.6	165.6	163.9
Total observations	214143	138443	149065
Unique reflections	9708 (962)	12775 (1264)	26997 (2651)
Multiplicity	22.3 (22.4)	10.8 (10.3)	5.4 (5.2)
Completeness (%)	99.9 (100.0)	99.9 (100.0)	99.8 (98.8)
$\langle I/\sigma(I)\rangle$	23.0 (8.3)	27.2 (3.6)	12.1 (2.6)
$R_{\rm merge}$ (%)	14.3 (82.0)	9.9 (82.8)	6.6 (68.7)
$R_{\rm meas}$† (%)	15.0 (85.8)	11.0 (91.1)	7.3 (75.7)
$R_{\rm meas0}$† (%)	15.3 (85.1)	11.7 (89.6)	ND

(*b*) Phasing (*AutoSol* in *PHENIX*).

Bragg spacing limits (Å)	Overall	11.29	7.08	5.51	4.67	4.12	3.73	3.43	3.19
Phasing power (r.m.s. anomalous $F_{\rm H}/E$), λ_1/λ_2	0.2/0.5	1.1/1.2	1.4/1.3	1.2/1.1	0.8/0.8	0.4/0.5	0.2/0.3	0.1/0.2	0.1/0.1
Phasing power (r.m.s. dispersive $F_{\rm H}/E$), λ_1 *versus* λ_2	0.3	0.8	1.0	0.8	0.6	0.4	0.2	0.2	0.2
Mean figure of merit	0.40	0.63	0.73	0.72	0.64	0.48	0.31	0.17	0.07
No. of Se sites per asymmetric unit (used/expected)	4/4								
Density-modification R (%)	24.7								

(*c*) Refinement (*REFMAC*5).

Resolution range (Å)	30–2.40 (2.46–2.40)
Reflections [$>0\sigma(F)$], working/test	24246/1350
R factor/$R_{\rm free}$ (%)	20.1/22.9
CC‡	1.000 (0.920)
R.m.s.d. bond lengths (Å)/angles (°)	0.011/1.518
No. of atoms	
Protein	2691
Sulfate ion	10
Water	70
Wilson B factor (Å^2)	56.22
Average B factor (Å^2)	
Protein	66.3
Sulfate ion	63.7
Water	60.5
Ramachandran plot (%)	
Most favoured	92.4
Additionally allowed	6.9
Generously allowed	0.7
Disallowed	0
Disordered regions (residue Nos.)	158–169, 387
PDB code	4fwv

† Multiplicity-weighted $R_{\rm merge}$ (Diederichs & Karplus, 1997). $R_{\rm meas}$ treats anomalous pairs as separate, whereas $R_{\rm meas0}$ assumes anomalous pairs to be equivalent. ‡ An estimation of the correlation of the averaged data set with the noise-free true signal (Karplus & Diederichs, 2012).

(TEV) protease-cleavage site (Glu-Asn-Leu-Tyr-Phe-Gln-Gly). Three mutations (Leu91Met, Leu188Met and Ile359Met) were introduced to facilitate *de novo* phasing. The N-terminal 6×His-tagged protein derivatized with selenomethionine was expressed in *Escherichia coli* B834(DE3) cells (Novagen) and grown in SelenoMet medium (Molecular Dimensions) following the manufacturer's instructions. After harvesting by centrifugation, the cells were resuspended in lysis/binding buffer (50 m*M* Tris–HCl pH 8.0, 500 m*M* NaCl) and ruptured using a high-pressure homogenizer (Avestin). After centrifugation, the supernatant was mixed with nickel-chelate affinity resin (Ni–NTA, Qiagen). The resin was washed with binding buffer containing 20 m*M* imidazole and the protein was eluted with binding buffer containing 250 m*M* imidazole. The protein was then subjected to TEV protease cleavage to remove the 6×His tag followed by buffer exchange and was reloaded onto an Ni–NTA column. After cleavage, the untagged protein contained an additional Gly-Gln dipeptide linker at the N-terminus. The flowthrough fraction containing the untagged MtaLonC-N was further purified using Mono Q and Superdex 75 columns (GE Healthcare). The protein was stored at 277 K in crystallization buffer consisting of 10 m*M* HEPES pH 7.5, 100 m*M* NaCl, 10%(*v*/*v*) glycerol, 1 m*M* DTT. The mutants MtaLonCδ and MtaLonCΔHHE were expressed and purified following the same procedure as used for the wild-type (WT) MtaLonC (Liao *et al.*, 2013).

2.2. Site-directed mutagenesis

The deletion mutations MtaLonCδ (removing residues 506–511) and MtaLonCΔHHE (replacing residues 118–205 with a triglycine linker) were generated by a PCR-based strategy using the QuikChange Kit (Stratagene) and the identities of the mutagenized products were verified by DNA sequencing.

2.3. Protein crystallization

Initial crystals of native or SeMet-derivatized MtaLonC-N were obtained by screening ∼400 conditions using a Phoenix RE crystallization robot (Rigaku, The Woodlands, Texas, USA). The crystallization condition was optimized at 295 K by varying the concentration of different precipitants, additives and buffers in around 50 two-dimensional grid optimization experiments using the sitting-drop vapour-diffusion technique. The best cube-like crystals of MtaLonC-N were grown by mixing 1 µl protein solution at 8 mg ml^{-1} with 1 µl well solution consisting of 400 m*M* lithium sulfate, 10 m*M* nickel

chloride, 100 m*M* Tris–HCl pH 8.4. Droplets were equilibrated against a reservoir containing 0.75 ml well solution. The crystals appeared in 7 d and reached dimensions of 0.2 × 0.2 × 0.2 mm in three months. Selected crystals were further dehydrated for 3 d in sitting drops equilibrated against 0.7–1.2 *M* lithium sulfate. All crystals were harvested and immersed briefly in mother liquor containing 30%(*v*/*v*) glycerol before data collection.

2.4. Structure determination

A peak data set to 3.4 Å resolution and a high-remote data set to 3.1 Å resolution were collected on beamline BL-1A of the Photon Factory from two different SeMet-substituted crystals with a mosaicity of about 0.5°. Totals of 720 and 360 images were collected for the peak and the high-remote data sets, respectively, with an oscillation angle of 0.5°, an exposure time of 1 s (peak) or 2 s (remote) per image and a crystal-to-detector distance of 339 mm (peak) or 345 mm (remote). All data were processed using the *HKL*-2000 package (Otwinowski & Minor, 1997; Table 1). The peak and remote data sets were truncated at resolutions of 3.4 and 3.1 Å resolution, respectively, before the $R_{\rm merge}$ (or $R_{\rm meas}$) value exceeded ∼0.9 (Table 1). However, in these resolution shells with poor merging statistics the empirical signal-to-noise ratio $I/\sigma(I)$ was still high, which was likely to be a consequence of crystal damage induced by synchrotron radiation. The structure was determined by multi-wavelength anomalous dispersion (MAD) using *AutoSol* in *PHENIX* (Adams *et al.*, 2010). All four expected selenium sites were identified with one molecule per asymmetric unit in space group *R*32. An initial model was built with *AutoBuild* in *PHENIX* based on the SeMet-phased density-modified electron-density map. A higher resolution data set to 2.4 Å resolution was subsequently collected on NSRRC beamline 13B1 (Taiwan) from a dehydrated native crystal with a mosaicity of around 0.4° and was used for subsequent model building using *Coot* (Emsley & Cowtan, 2004) and refinement using *REFMAC*5 (Murshudov *et al.*, 2011). For the 2.4 Å resolution data set a total of 180 images were collected with an oscillation angle of 0.5°, an exposure time of 5 s per image and a crystal-to-detector distance of 300 mm. Crystallographic and refinement statistics are listed in Table 1. The N-terminal dipeptide linker from the vector, loop residues 158–169 and the C-terminal Ser387 were not modelled owing to a lack of electron density for these residues. All structure figures were prepared with *PyMOL* (v.1.3; Schrödinger).

2.5. Molecular modelling of full-length MtaLonC

To create a spliced model of MtaLonC hexamers including the six helical hairpin extensions (HHEs), the structure of MtaLonC-N was superimposed onto that of the apo MtaLonC monomer (PDB entry 4fw9; Liao *et al.*, 2013) with reference to their consensus helix h4 (residue 226–240). Therefore, the spliced MtaLonC monomer took the atoms of helices h1, h2 and h3 from the MtaLonC-N structure, which were missing in apo MtaLonC, and all of the other atoms from apo MtaLonC. The 12 missing residues (residues 158–169) in the reverse loop of the HHE were rebuilt using *MODELLER* v.9.10 (Eswar *et al.*, 2007) using the spliced MtaLonC monomer as the template. The loop configuration was refined by minimizing a pseudo-energy function in *MODELLER*, which was set up to allow minor conformational changes at residues 149–157 and 170–175. Next, the hexameric assembly was obtained by applying a sixfold symmetry operation to the modelled monomer. There were no atomic clashes in the assembled hexameric structure except that a subset of atoms in the modelled missing loop came too close to those in helix h2 of the HHE from the neighbouring monomer, with the shortest interatomic distance between the two being 0.2 Å. Energy-minimization and molecular dynamics simulations were then carried out using *NAMD* v.2.9 (Phillips *et al.*, 2005) with the CHARMM22 force field (MacKerell *et al.*, 1998) to remove the clashes and to energetically relax the modelled loop 158–169 in the HHE while constraining the rest of the spliced structures, including h1–h4 of the HHE. The energy minimization was carried out in explicit water for 100 ps and the spliced hexamer was simulated at 310 K and 100 kPa for 10 ns, restraining the heavy atoms that were not in the modelled loop 158–169 with a force constant of 500 kcal mol^{-1} Å^{-2} (1 cal = 4.184 J).

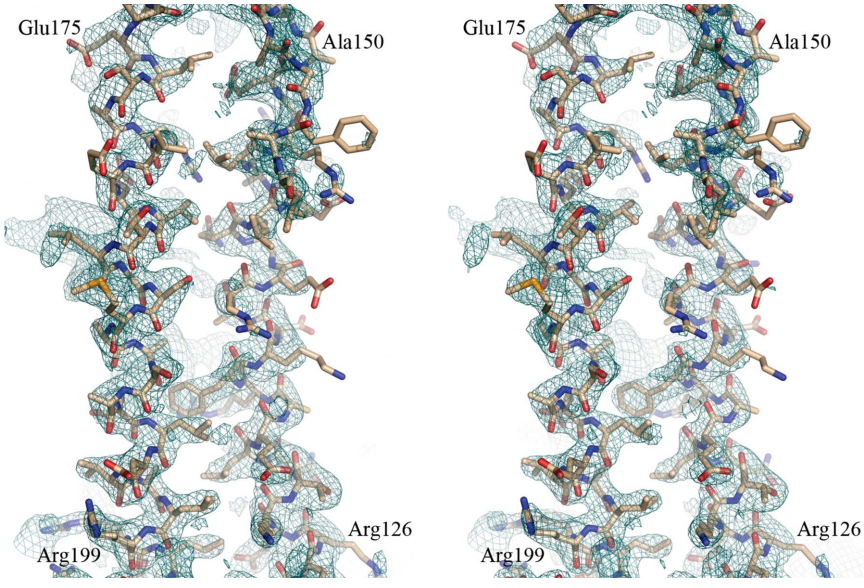

Figure 2
Stereoview of an electron-density map ($2F_{\rm o} - F_{\rm c}$) covering part of the HHE region showing residues 126–150 and 175–199 and contoured at 1.0σ.

2.6. Proteolytic activity assays

A degradation assay of bovine casein (Sigma, USA), purified by Mono Q chromatography, by WT and mutants of MtaLonC was carried out by incubating the

substrates (300 µg) with the enzyme (30 µg) in a buffer consisting of 50 m*M* Tris–HCl, 1 m*M* DTT, 10 m*M* Na_2HPO_4 pH 8.0 for the indicated time periods at 328 K. The reaction products were analyzed by SDS–PAGE. A fluorescence peptide assay using the synthetic fluorogenic peptide F-β20-Q was performed as described previously (Liao *et al.*, 2013).

2.7. Lysozyme-aggregation assay

The assay was performed as described by Peschek *et al.* (2009). In brief, chicken egg-white lysozyme (Sigma–Aldrich) was dissolved in PBS buffer pH 7.4 (Sigma–Aldrich) and then diluted into denaturing buffer in PBS with 1 m*M* tris(2-carboxyethyl)phosphine (TCEP; Sigma–Aldrich) only at 12.5 µ*M* or with various concentrations of the MtaLonC mutants. The aggregation of reduced denatured lysozyme was monitored at 360 nm for 30 min at 310 K.

3. Results and discussion

3.1. Overall structure of MtaLonC-N (residues 35–387)

We have recently determined the crystal structure of a LonC protease from *M. taiwanensis* (MtaLonC). Despite the integrity of the full-length protein in the crystal, a significant portion of the LID was disordered and only diffuse and broken electron density was associated with it (Liao *et al.*, 2013). In order to

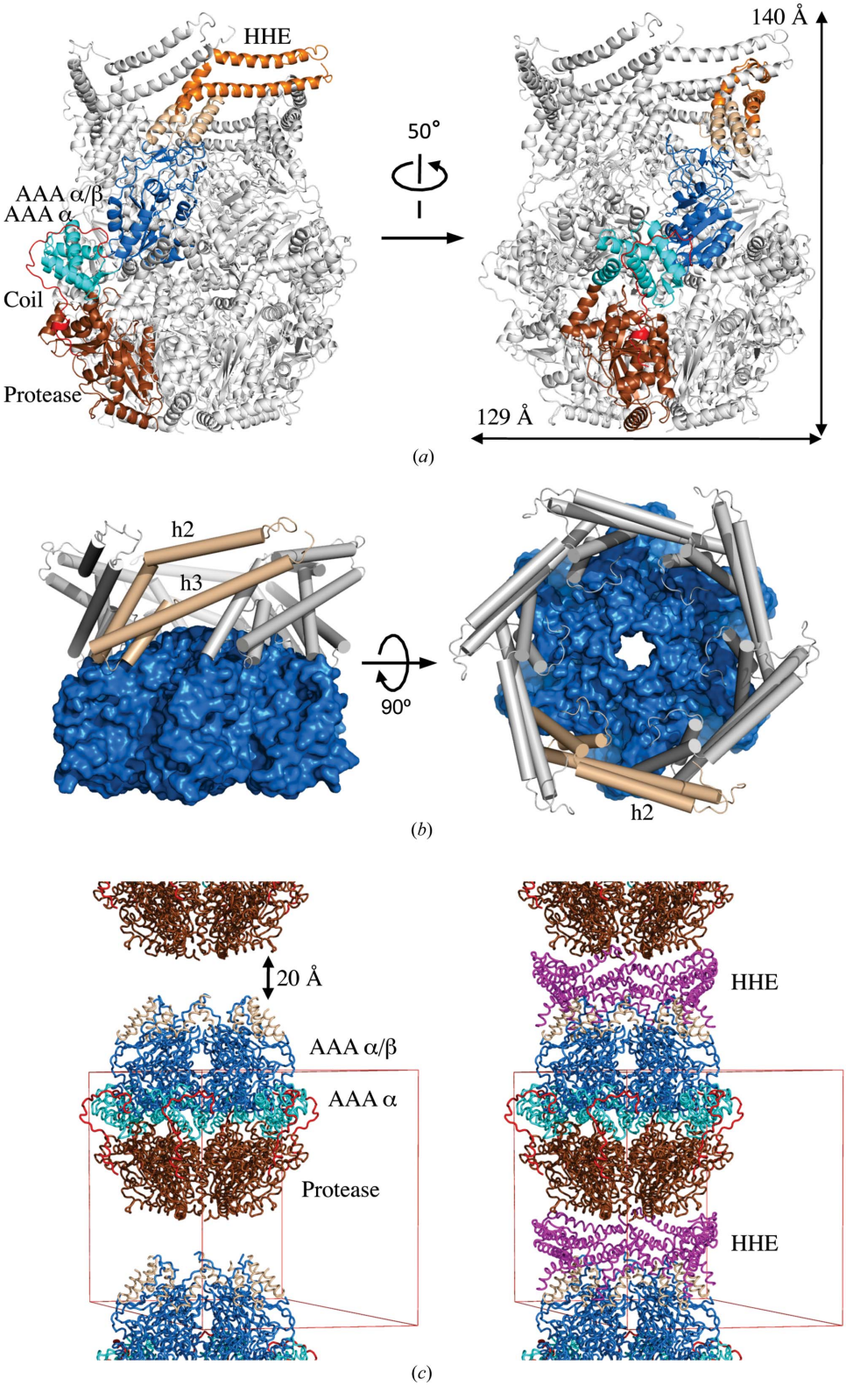

Figure 3
Spliced models of full-length MtaLonC. (*a*) Two side views of the ribbon diagram of MtaLonC in its hexameric assembly, with one monomer coloured using a similar scheme as in Fig. 1(*a*) to highlight its highly twisted and extended conformation. The portion of the HHE domain that may be structurally flexible (see text) is coloured orange. Also indicated are the particle dimensions. (*b*) Two orthogonal views of spliced HHEs as cylindrical helices, with one monomer highlighted in wheat, on top of the upper portion of the hexameric chamber of MtaLonC in surface representation. In this diagram the AAA α subdomain and the protease domain are removed for clarity. The open entry portal is evident in the top view (right). The HHE domains are modelled as described in the text. (*c*) A lateral view of the molecular packing in the MtaLonC crystal, as reported previously (Liao *et al.*, 2013), is shown on the left. The same lateral view of the molecular packing on the right shows the good fitting of the HHEs (coloured magenta), based on the structure of MtaLonC-N (this work), into the 20 Å gap between the barrel-like particles of MtaLonC. The *c* axis is vertical; the *ab* plane is horizontal and perpendicular to the plane of the page. MtaLonC is rendered in tubes using a similar colouring scheme as in Fig. 1(*a*).

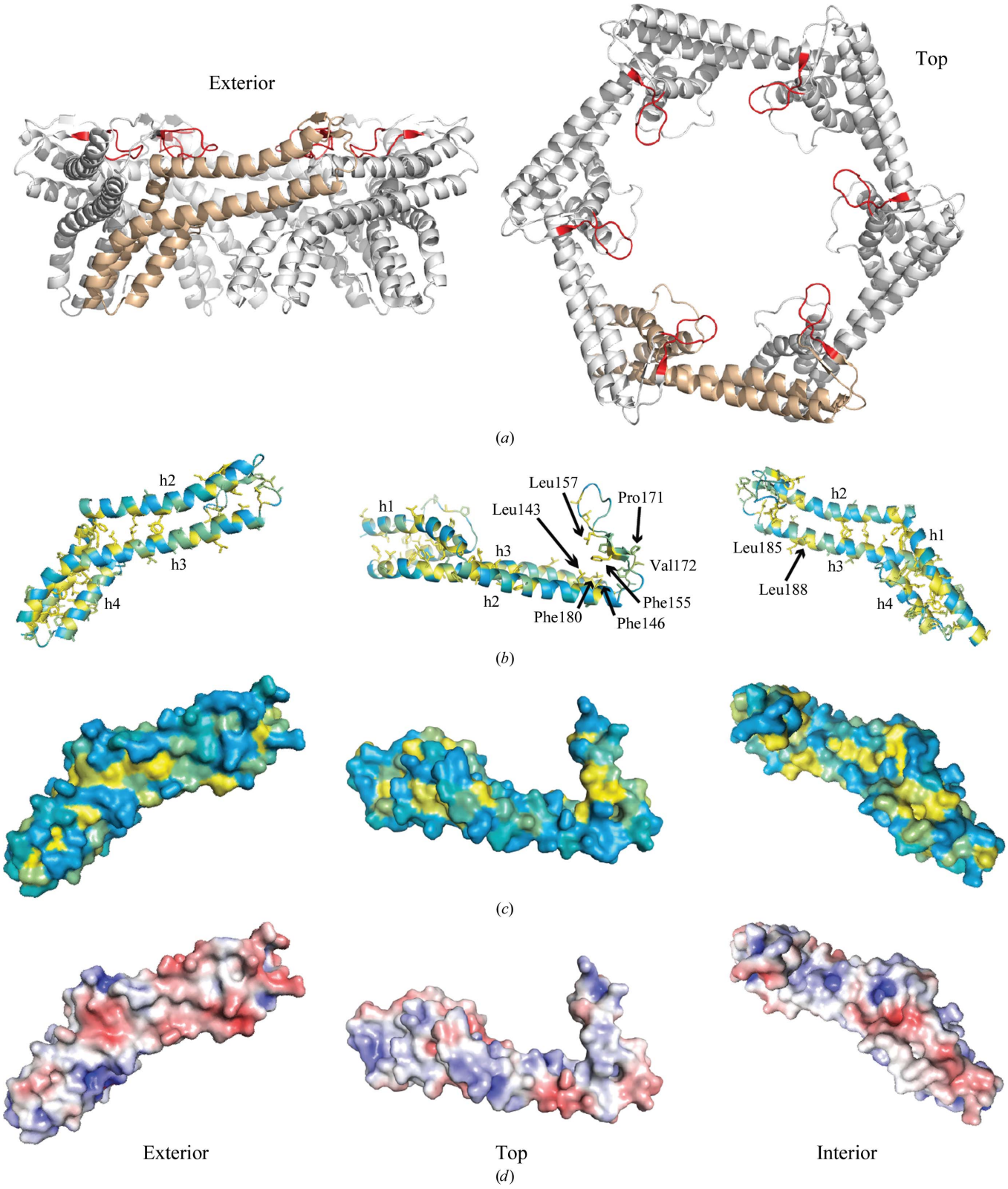

Figure 4
Molecular surface of the HHE (residues 97–252). (*a*) Two orthogonal views of HHEs in a hexameric assembly with one monomer highlighted in wheat. The tip region (residues 158–169) of the reverse loop, coloured in red, was modelled as described in §2. (*b*) Views of one HHE monomer from the exterior, the top and the interior are shown in ribbon representation. (*c*) Views of the HHE in the same orientations as in (*b*) in hydrophobic polar surface representation. The colouring in (*b*) and (*c*) runs approximately from aqua (−3.0; charged or polar) to yellow (+1.0; hydrophobic) according to a previously defined hydrophobicity scale for amino acids (Eisenberg *et al.*, 1984). (*d*) The same three views of the HHE in electrostatic surface representation. The surface is coloured according to its electrostatic potential from red ($-5kT\ e^{-1}$; negatively charged) to blue ($+5kT\ e^{-1}$; positively charged). Note that the electrostatic potential mapped to the surface of the missing tip region (residues 158–169) of the reverse loop is purely speculative.

resolve the structure of the disordered domain, we designed a truncated construct (MtaLonC-N; residues 35–387) encompassing the AAA+ α/β-subdomain and the LID (Fig. 1*a*). The crystal structure of MtaLonC-N was determined by two-wavelength anomalous dispersion and the model was refined to 2.4 Å resolution (R_{work} and R_{free} of 20.1 and 22.9%, respectively; Table 1). The LID of MtaLonC emerges from the AAA+ α/β core as an AAA+ insert known as Ins1 (Iyer *et al.*, 2004), which forms a base consisting of a three-stranded β-sheet (Fig. 1*b*). Fused to the base is a large α-helical hairpin extension (hereafter termed HHE; residues 97–252), in which ~50 residues extend out as a series of two arched α-helical segments (h1–h2) followed by a 25-residue loop. Although 12 residues (amino acids 158–169) are missing from the electron density, this loop has a shape like a hook. In the coiled-coil topology the polypeptide chain loosely packs against h1 and h2 as a long curved helical segment (h3) followed by a straight helix (h4) such that h1, h3 and h4 form a three-helical bundle which connects to the base *via* an extended coil (Fig. 1*b*). Notably, many of the side chains of the neighbouring h2–h3 helices are flexible and the two helices do not engage in canonical leucine zipper-like rigid packing (Fig. 2). The structure suggests the ability of the six HHEs to adopt different conformations and to occupy different positions in the hexameric assembly. Previously, in the structure of the full-length MtaLonC only the beginning portion of helix h1 and the whole helix h4 with the C-terminal coil region of the LID were resolved, which had weak electron density associated with high *B*-factor values (Liao *et al.*, 2013). Here, the HHE helical regions resolved in the structure of MtaLonC-N are not involved in intermolecular contacts in the *R*32 unit cell, suggesting that the helical hairpin structure is not induced by crystal-packing interactions.

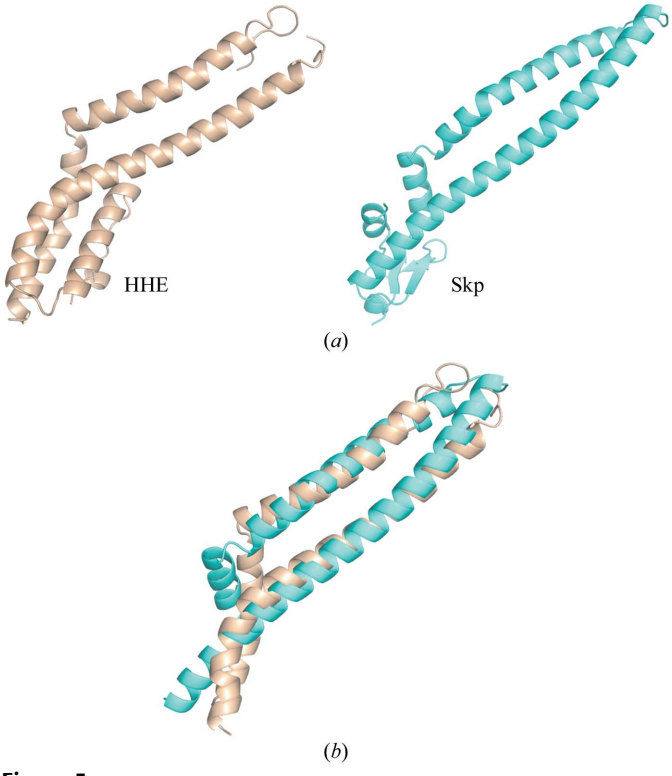

Figure 5
Structural similarity of the HHE to the α-helical tentacles of cytosolic chaperones. (*a*) Ribbon representations of the HHE of MtaLonC (left) and the Skp monomer (right). (*b*) Superimposed helical hairpin structures of the HHE of MtaLonC (wheat) and Skp (cyan) shown in ribbon representation.

3.2. Quaternary structure based on a spliced model

To obtain insight into the architectural features of the MtaLonC hexamer including intact LIDs, we have built a complete model of MtaLonC by superimposing helix h4 of MtaLonC-N onto that of the crystal structure of full-length MtaLonC (PDB entry 4fw9; r.m.s.d. = 0.49 Å) and splicing the HHE into the latter as described in §2. In the spliced model, the six tentacle-like HHEs protrude from the top surface of the AAA+ ring, forming a fence-shaped basket over the top pore (Figs. 3*a* and 3*b*). The six HHEs of the spliced model generated in this way do not have any steric hindrance between adjacent subunits. Previously, in the crystal structure of full-length MtaLonC each hexameric barrel packed into stable planar layers perpendicular to the crystallographic sixfold axis *via* side-by-side interactions; surprisingly, a 20 Å gap was observed between the planar layers of the MtaLonC barrels (Liao *et al.*, 2013). As shown in Fig. 3(*c*) (left), the unresolved HHEs would occupy the substantial gap between the upper AAA+ ring of a barrel and the bottom protease ring of the barrel from the neighbouring layer on top of it. We find that the spliced HHE model can now be placed into the lattice of the previous crystal of full-length MtaLonC with no clashes between symmetry-related monomers and with the HHEs fitting snugly into the gap between the layers perpendicular to the *c* axis (Fig. 3*c*; right).

In the spliced model the reverse loops from the six HHEs are located at the top edge of the basket. We have modelled the 12 missing residues constituting the tip region of the loop and analyzed the surface properties of the HHE (Fig. 4*a*). A majority of the hydrophobic residues are located at the h1–h3 and h2–h3 helical interfaces and are particularly found on the interior side, at the junction of the h2–h3 coil and the reverse loop (Figs. 4*b*, 4*c* and 4*d*). Specifically, Leu143, Leu146, Leu180, Phe155, Leu157, Val172, Leu185 and Leu188 contribute substantial exposed hydrophobic areas in the interior of the basket (Figs. 4*a* and 4*b*). The disordered loop residues consist of both hydrophobic and charged residues (Supplementary Fig. S1[1]). Most of the hydrophobic residues of the HHE are conserved in LonC proteases (Supplementary Fig. S1). A hook-like structure with hydrophobic interior may be suitable for trapping unfolded protein substrates while assuming structural malleability. Although the HHEs are modelled here with sixfold symmetry, they might assume different conformations and may emanate from the AAA+ ring at different angles in solution; the HHE in the spliced model may display only one of many possible conformers.

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: WD5213). Services for accessing this material are described at the back of the journal.

Interestingly, based on its structural features, the non-homologous N-terminal substrate-recognition domain of *E. coli* LonA has previously been proposed to exhibit an independent horizontal sweeping motion (Li *et al.*, 2010).

3.3. Structural and functional similarity to cytosolic chaperones

A search using the *DALI* server (Holm & Rosenström, 2010) against h1–h4 of HHE revealed its marked structural similarity to the 'seventeen kilodalton protein' (Skp; *Z*-score = 6.9; r.m.s.d. = 3.8 Å over 143 residues with 5% sequence identity; Fig. 5*a*). Skp is a periplasmic chaperone that is known to protect unfolded substrates from aggregation. It has a molecular shape consisting of three helical hairpin tentacles protruding from a trimeric β-barrel core, defining a basket-like structure that binds unfolded protein substrates (Korndörfer *et al.*, 2004; Walton & Sousa, 2004). The overall shape of Skp is similar to another cytosolic chaperone, prefoldin, which has a basket-like structure with six paired coiled coils (Siegert *et al.*, 2000). Both the HHE of MtaLonC and the helical hairpin tentacle of Skp consist of a series of two arched/kinked helices that are loosely packed against one long curved helix (Figs. 5*a* and 5*b*). Previously, chaperone activity has been identified in LonA proteases, which can inhibit the aggregation of denatured model substrates independently of the protease domain (Lee *et al.*, 2004; Bartoszewska *et al.*, 2012). To determine whether MtaLonC has chaperone activity, we constructed MtaLonCδ, a mutant with no protease activity owing to the removal of the β-hairpin (Glu506–Trp511) in the substrate-binding groove (Liao *et al.*, 2013). Furthermore, an internally truncated mutant MtaLonCΔHHE was constructed in which helix h2, the reverse loop and most of helix h3 (residues 118–205) were removed and replaced by a triglycine linker. We found that MtaLonCδ suppresses the aggregation of denatured lysozyme in a dosage-dependent manner, an activity that is also exhibited by Skp (Walton & Sousa, 2004; Fig. 6*a*). In contrast, MtaLonCΔHHE has not only lost the activity to inhibit lysozyme aggregation, but the mutant even causes exacerbated aggregation of the denatured substrate (Fig. 6*b*). Such contrasting activity has also been observed in similar mutational analyses of other chaperones (Giese *et al.*, 2005; Liao *et al.*, 2009). Moreover, unlike full-length MtaLonC, MtaLonCΔHHE cannot degrade the exogenous protein substrate casein and has significantly reduced cleavage activity towards the fluorogenic peptide substrate F-β20-Q (Gur & Sauer, 2008; Liao *et al.*, 2012; Figs. 6*c* and 6*d*). These results suggest that the six HHEs of MtaLonC are involved in recognition of the unfolded protein substrates. Therefore, the LID of LonC may serve a dual function both as an Skp-like chaperone to prevent protein aggregation and as a barbed basket to collect protein substrates and facilitate their access to the degradation chamber through the AAA+ pore. Finally, it is worth noting that the HHE domain of LonC is reminiscent of the coiled-coil middle domain (MD) of Hsp104/ClpB. The leucine-rich MD of TClpB, a bacterial Hsp104 homologue, forms a structure with the shape of a two-bladed propeller which was suggested to act as a molecular 'crowbar' that pries apart large aggregates when coupled to ATP hydrolysis by the tandem AAA+ modules (Lee *et al.*, 2003). MtaLonC lacks intrinsic ATPase activity; our structural and functional analyses revealed that the HHE domain forms a coiled coil-like structure and has an ATP-independent chaperone activity to prevent the aggregation of a model substrate. Interestingly, the MD domain is similarly inserted into the AAA+ module of TClpB; however, it emerges from the small α subdomain rather than the large α/β subdomain as is the case for the HHE domain of LonC.

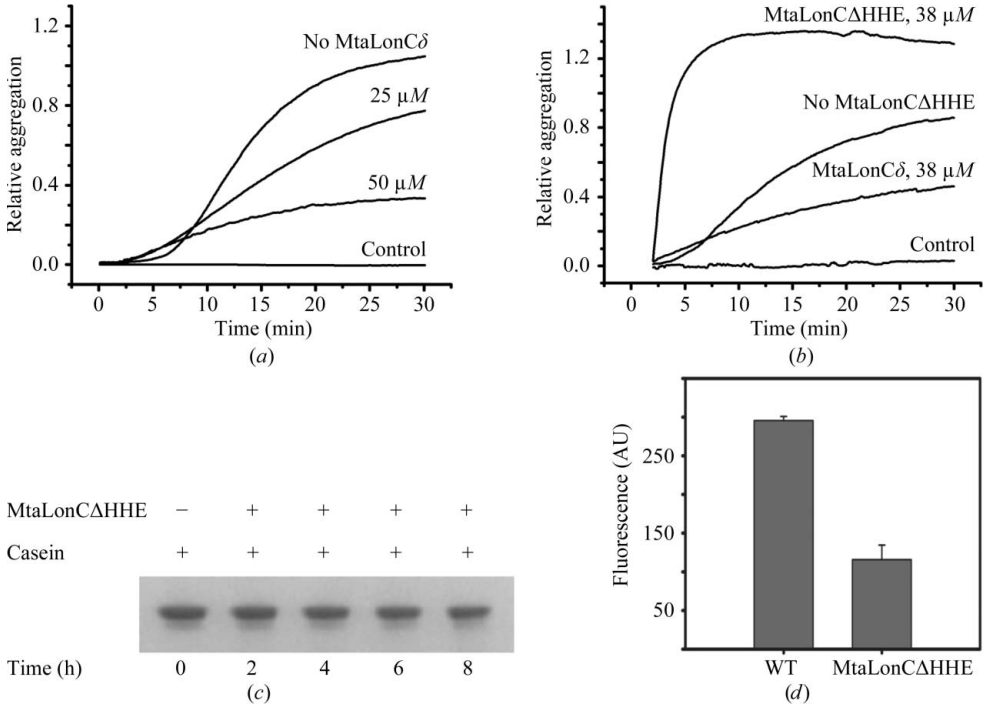

Figure 6
Functional activity of the HHEs. (*a*) Inhibition of the aggregation of denatured lysozyme by a proteolytically deficient MtaLonC mutant (MtaLonCδ; see text). (*b*) Lack of aggregation inhibition by HHE-deleted MtaLonC (MtaLonCΔHHE). Controls are the respective MtaLonC enzymes alone incubated with TCEP. (*c*) Proteolytic degradation of α_{S2}-casein by MtaLonCΔHHE. (*d*) Comparison of the proteolytic cleavage of the fluorogenic peptide F-β20-Q by wild-type (WT) MtaLonC and MtaLonCΔHHE. The increased fluorescence upon peptide cleavage by the enzymes was measured after a 30 min incubation. The error bars show means with standard deviations (n = 3).

In conclusion, the present work describes the complete structure of the LID of MtaLonC. In the structure of full-length MtaLonC

most of the LID was crystallographically disordered. Therefore, the structural results reported here may suggest a general architectural feature of Lon, which includes a hexameric degradation chamber and a basket on top of the chamber defined by six copies of the conformationally flexible LID. In LonC, the LID assumes a tentacle-like helical hairpin extension with a flexible reverse loop. In archaeal LonB, the LID is composed of a series of transmembrane segments. Although the quaternary structure of these membrane-anchoring helices remains unknown, it is likely that they form a membrane-spanning channel atop the LonB chamber. The LID/LAN of LonA is known to contain an N-terminal globular domain and a long helix. Interestingly, the C-terminus of this long helix is known to contain several trypsin-sensitive sites (Duman & Löwe, 2010; Li *et al.*, 2010; Roudiak & Shrader, 1998; Vasilyeva *et al.*, 2002), indicating the existence of a flexible joint in the LID of LonA. Recently, this region has been shown to be critical for interaction with certain protein substrates (Cheng *et al.*, 2012). Further biophysical analysis of an intact LID of LonA would be necessary to determine whether this structurally unknown trypsin-sensitive region forms a LonC-like reverse loop.

We thank Chih-I Chang for editing the manuscript. We are grateful to Yi-Hui Chen at the 13B1 beamline of the National Synchrotron Radiation Research Center (NSRRC), Taiwan and the staff at the Photon Factory, Tsukuba, Japan for their support during synchrotron data collection. We acknowledge the Core Facilities for Protein Structural Analysis at Academia Sinica and the National Center for High-performance Computing for providing facilities and computer time for this work. This work was supported in part by Academia Sinica
(to C-IC) and by grants NSC101-2311-B-001-007 (to J-HL), NSC100-2113-M-001-022-MY3 (to S-HW), NSC101-2627-M-007-010 (to L-WY) and NSC101-2811-M-001-164 (to C-IC).

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Bartoszewska, M., Williams, C., Kikhney, A., Opaliński, Ł., van Roermund, C. W., de Boer, R., Veenhuis, M. & van der Klei, I. J. (2012). *J. Biol. Chem.* **287**, 27380–27395.
Cha, S.-S., An, Y. J., Lee, C. R., Lee, H. S., Kim, Y.-G., Kim, S. J., Kwon, K. K., De Donatis, G. M., Lee, J.-H., Maurizi, M. R. & Kang, S. G. (2010). *EMBO J.* **29**, 3520–3530.
Cheng, I., Mikita, N., Fishovitz, J., Frase, H., Wintrode, P. & Lee, I. (2012). *J. Mol. Biol.* **418**, 208–225.
Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Mol. Biol.* **4**, 269–275.
Duman, R. E. & Löwe, J. (2010). *J. Mol. Biol.* **401**, 653–670.
Ebel, W., Skinner, M. M., Dierksen, K. P., Scott, J. M. & Trempy, J. E. (1999). *J. Bacteriol.* **181**, 2236–2243.
Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. (1984). *J. Mol. Biol.* **179**, 125–142.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M.-Y., Pieper, U. & Sali, A. (2007). *Curr. Protoc. Protein Sci.*, Unit 2.9. doi:10.1002/0471140864.ps0209s50.
Giese, K. C., Basha, E., Catague, B. Y. & Vierling, E. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 18896–18901.
Goldberg, A. L., Moerschell, R. P., Chung, C. H. & Maurizi, M. R. (1994). *Methods Enzymol.* **244**, 350–375.
Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Granot, Z., Melamed-Book, N., Bahat, A. & Orly, J. (2007). *Mol. Cell. Endocrinol.* **265–266**, 51–58.
Gur, E. (2013). *Subcell. Biochem.* **66**, 35–51.
Gur, E. & Sauer, R. T. (2008). *Genes Dev.* **22**, 2267–2277.
Gur, E., Vishkautzan, M. & Sauer, R. T. (2012). *Protein Sci.* **21**, 268–278.
Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
Iyer, L. M., Leipe, D. D., Koonin, E. V. & Aravind, L. (2004). *J. Struct. Biol.* **146**, 11–31.
Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
Korndörfer, I. P., Dommel, M. K. & Skerra, A. (2004). *Nature Struct. Mol. Biol.* **11**, 1015–1020.
Lee, A. Y.-H., Hsu, C.-H. & Wu, S.-H. (2004). *J. Biol. Chem.* **279**, 34903–34912.
Lee, S., Sowa, M. E., Watanabe, Y. H., Sigler, P. B., Chiu, W., Yoshida, M. & Tsai, F. T. (2003). *Cell*, **115**, 229–240.
Li, M., Gustchina, A., Rasulova, F. S., Melnikov, E. E., Maurizi, M. R., Rotanova, T. V., Dauter, Z. & Wlodawer, A. (2010). *Acta Cryst.* D**66**, 865–873.
Li, M., Rasulova, F., Melnikov, E. E., Rotanova, T. V., Gustchina, A., Maurizi, M. R. & Wlodawer, A. (2005). *Protein Sci.* **14**, 2895–2900.
Liao, J.-H., Ihara, K., Kuo, C.-I., Huang, K.-F., Wakatsuki, S., Wu, S.-H. & Chang, C.-I. (2013). *Acta Cryst.* D**69**, 1395–1402.
Liao, J.-H., Kuo, C.-I., Huang, Y.-Y., Lin, Y.-C., Lin, Y.-C., Yang, C.-Y., Wu, W.-L., Chang, W.-H., Liaw, Y.-C., Lin, L.-H., Chang, C.-I. & Wu, S.-H. (2012). *PLoS One*, **7**, e40226.
Liao, J.-H., Lee, J.-S., Wu, S.-H. & Chiou, S.-H. (2009). *Mol. Vis.* **15**, 1429–1444.
MacKerell, A. D. *et al.* (1998). *J. Phys. Chem. B*, **102**, 3586–3616.
Mizusawa, S. & Gottesman, S. (1983). *Proc. Natl Acad. Sci. USA*, **80**, 358–362.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Notredame, C., Higgins, D. G. & Heringa, J. (2000). *J. Mol. Biol.* **302**, 205–217.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Peschek, J., Braun, N., Franzmann, T. M., Georgalis, Y., Haslbeck, M., Weinkauf, S. & Buchner, J. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 13272–13277.
Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L. & Schulten, K. (2005). *J. Comput. Chem.* **26**, 1781–1802.
Rotanova, T. V., Botos, I., Melnikov, E. E., Rasulova, F., Gustchina, A., Maurizi, M. R. & Wlodawer, A. (2006). *Protein Sci.* **15**, 1815–1828.
Roudiak, S. G. & Shrader, T. E. (1998). *Biochemistry*, **37**, 11255–11263.
Sauer, R. T. & Baker, T. A. (2011). *Annu. Rev. Biochem.* **80**, 587–612.
Selkoe, D. J. (2003). *Nature (London)*, **426**, 900–904.
Siegert, R., Leroux, M. R., Scheufler, C., Hartl, F. U. & Moarefi, I. (2000). *Cell*, **103**, 621–632.
Vasilyeva, O. V., Kolygo, K. B., Leonova, Y. F., Potapenko, N. A. & Ovchinnikova, T. V. (2002). *FEBS Lett.* **526**, 66–70.
Walton, T. A. & Sousa, M. C. (2004). *Mol. Cell*, **15**, 367–374.
Wickner, S., Maurizi, M. R. & Gottesman, S. (1999). *Science*, **286**, 1888–1893.